Supplementary Document for Generalized Fiducial Inference: A Review and New Results

Jan Hannig Hari Iyer Randy C. S. Lai Thomas C. M. Lee

August 26, 2016

Abstract

This is a supplementary document for the paper Generalized Fiducial Inference: A Review and New Results.

A Proof of Theorem 1

Recall the data generating equation (1), and assume that $\boldsymbol{U} \in \mathbb{R}^n$ is an absolutely continuous random vector with a joint density $f_{\boldsymbol{U}}(\boldsymbol{u})$, defined with respect to the Lebesgue measure on $\mathbb{R}^n$, continuous on its support $\mathcal{U}$. We need the following assumptions.

Assumption A.1. The function $\boldsymbol{G}$ has continuous partial derivatives with respect to all variables $\theta_j,\ j = 1, \ldots, p$ and $u_i,\ i = 1, \ldots n$.

Assumption A.2. For each $\boldsymbol{y}$ and $\boldsymbol{\theta}$ there is at most one $\boldsymbol{u} \in \mathcal{U}$ so that $\boldsymbol{y} = \boldsymbol{G}(\boldsymbol{u}, \boldsymbol{\theta})$. For the observed data $\boldsymbol{y}$ there is a $\boldsymbol{\theta}$ and $\boldsymbol{u} \in \mathcal{U}$ so that $\boldsymbol{y} = \boldsymbol{G}(\boldsymbol{u}, \boldsymbol{\theta})$. Additionally, the determinant of the $n \times n$ Jacobian matrix

$$\det\left(\frac{\boldsymbol{d}}{\boldsymbol{du}}\boldsymbol{G}(\boldsymbol{u}, \boldsymbol{\theta})\right) \neq 0$$

for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and $\boldsymbol{u} \in \mathcal{U}$.

Assumption A.3. The $n \times p$ Jacobian matrix $\frac{\boldsymbol{d}}{\boldsymbol{d\theta}}G(\boldsymbol{U}, \boldsymbol{\theta})$ is of rank p.

For Part (iii) of Theorem 1 we will also need the following assumption.

Assumption A.4. The entries of the Jacobian matrix $\frac{\boldsymbol{d}}{\boldsymbol{d\theta}}\boldsymbol{G}(\boldsymbol{u},\boldsymbol{\theta})$ have continuous partial derivatives with respect to all variables $\theta_j,\ j=1,\dots,p$ and $u_i,\ i=1,\dots n$.

The proof of Theorem 1 begins here. We first derive a useful formula for the likelihood function $f(\boldsymbol{y}|\boldsymbol{\theta})$. Consider the implicit function

$$\boldsymbol{y}-\boldsymbol{G}(\boldsymbol{u},\boldsymbol{\theta})=0. \tag{1}$$

If for a fixed $\boldsymbol{y}$ and $\boldsymbol{\theta}$ there is $\boldsymbol{u}$ solving (1), the implicit function theorem using Assumptions A.1 and A.2 implies that there is a neighborhood of $(\boldsymbol{y},\boldsymbol{u})$ on which the function $\boldsymbol{u}(\boldsymbol{y})$ is uniquely defined. Moreover the function $\boldsymbol{u}(\boldsymbol{y})$ is continuously differentiable and simple calculation shows that on this neighborhood the Jacobian matrix

$$\frac{\boldsymbol{du}(\boldsymbol{y})}{\boldsymbol{dy}}=\left(\frac{\boldsymbol{d}}{\boldsymbol{du}}\boldsymbol{G}(\boldsymbol{u},\boldsymbol{\theta})\right)^{-1}\Bigg|_{\boldsymbol{u}=\boldsymbol{G}^{-1}(\boldsymbol{y},\boldsymbol{\theta})}.$$

Consequently, since by Jacobian transformation theorem

$$f(\boldsymbol{y}|\boldsymbol{\theta})=f_{\boldsymbol{U}}(\boldsymbol{G}^{-1}(\boldsymbol{y},\boldsymbol{\theta}))\left|\det\left(\frac{\boldsymbol{du}(\boldsymbol{y})}{\boldsymbol{dy}}\right)\right|,$$

On the other hand, if for a fixed $\boldsymbol{y}$ there is no solution $\boldsymbol{u}$ then $f(\boldsymbol{y}|\boldsymbol{\theta})=0$ by definition. In any case

$$f(\boldsymbol{y}|\boldsymbol{\theta})=\frac{f_{\boldsymbol{U}}(\boldsymbol{u})}{\left|\det\left(\frac{\boldsymbol{d}}{\boldsymbol{du}}\boldsymbol{G}(\boldsymbol{u},\boldsymbol{\theta})\right)\right|}\Bigg|_{\boldsymbol{u}=\boldsymbol{G}^{-1}(\boldsymbol{y},\boldsymbol{\theta})}.$$

Part (i): This is a special case of Part (ii).

Part (ii): For each $1\le i_1<\cdots<i_p\le n$ define a multi-index $\boldsymbol{i}=\{i_1,\dots,i_p\}$ and a vector $\boldsymbol{y}_{\boldsymbol{i}}=(y_{i_1},\dots,y_{i_p})$. Next define the complement multi-index $\boldsymbol{i}^{\complement}=\{i,i\notin\boldsymbol{i}\}$ and its corresponding vector $\boldsymbol{y}_{\boldsymbol{i}^{\complement}}$. Let us now consider the implicit function $\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y}_{\boldsymbol{i}^{\complement}})$ defined by (1) with $\boldsymbol{y}_{\boldsymbol{i}}$ held fixed at the observed values.

Fix $\boldsymbol{u}=\boldsymbol{G}^{-1}(\boldsymbol{y},\boldsymbol{\theta})$. If the determinant of the $p\times p$ matrix obtained by keeping only rows $\boldsymbol{i}=(i_1,\dots,i_p)$ of the $n\times p$ Jacobian matrix

$$\det\left(\frac{\boldsymbol{d}}{\boldsymbol{d\theta}}\boldsymbol{G}(\boldsymbol{u},\boldsymbol{\theta})\right)_{\boldsymbol{i}}\neq 0, \tag{2}$$

then a direct use of implicit function theorem shows that the $n \times n$ Jacobian matrix

$$\frac{\boldsymbol{d}\boldsymbol{u}(\boldsymbol{\theta}, \boldsymbol{y}_{\boldsymbol{i}^{\complement}})}{\boldsymbol{d}\boldsymbol{\theta}\boldsymbol{y}_{\boldsymbol{i}^{\complement}}} = \left(\frac{\boldsymbol{d}}{\boldsymbol{d}\boldsymbol{u}}\boldsymbol{G}(\boldsymbol{u}, \boldsymbol{\theta})\right)^{-1} \left(\frac{\boldsymbol{d}}{\boldsymbol{d}\boldsymbol{\theta}}\boldsymbol{G}(\boldsymbol{u}, \boldsymbol{\theta}) \ , \ \frac{\boldsymbol{d}\boldsymbol{y}}{\boldsymbol{d}\boldsymbol{y}_{\boldsymbol{i}^{\complement}}}\right)\Bigg|_{\boldsymbol{u}=\boldsymbol{G}^{-1}(\boldsymbol{y},\boldsymbol{\theta})},$$

where the last matrix is obtained by concatenating the columns of the $n \times p$ and $n \times (n-p)$ Jacobian matrices on either side of the vertical line. Consequently, the joint density of the random vector $(\boldsymbol{\theta}, \boldsymbol{Y}_{\boldsymbol{i}^{\complement}})$ evaluated at the observed value $\boldsymbol{y}_{\boldsymbol{i}^{\complement}}$ is

$$h_{\boldsymbol{i}}(\boldsymbol{\theta}, \boldsymbol{y}) = f(\boldsymbol{y}|\boldsymbol{\theta}) \left| \det \left(\frac{\boldsymbol{d}}{\boldsymbol{d}\boldsymbol{\theta}}\boldsymbol{G}(\boldsymbol{u}, \boldsymbol{\theta})\right)_{\boldsymbol{i}} \right|.$$

Hannig (2013) shows that the fiducial density can be computed as proportional to the sum of the joint densities

$$r(\boldsymbol{\theta}|\boldsymbol{y}) \propto \sum_{\boldsymbol{i}=(i_1,\ldots,i_p)} h_{\boldsymbol{i}}(\boldsymbol{\theta}, \boldsymbol{y}),$$

taken as a function of $\boldsymbol{\theta}$ with $\boldsymbol{y}$ fixed at the observed values. There is a caveat that if for some $\boldsymbol{i}$ (2) is not satisfied, the term corresponding to that $\boldsymbol{i}$ is missing from the sum as it is of a lower order in the calculation of the fiducial density. Assumption A.3 guarantees that there is at least one term not missing and the formula is still formally true with zeros substituted for the missing terms. The statement of Part (ii) follows.

Part (iii) Again, fix the value $\boldsymbol{u} = \boldsymbol{G}^{-1}(\boldsymbol{y}, \boldsymbol{\theta})$. Consider the singular value decomposition

$$\frac{\boldsymbol{d}}{\boldsymbol{d}\boldsymbol{\theta}}\boldsymbol{G}(\boldsymbol{u}, \boldsymbol{\theta}) = A(\boldsymbol{u}, \boldsymbol{\theta})S(\boldsymbol{u}, \boldsymbol{\theta})B(\boldsymbol{u}, \boldsymbol{\theta}).$$

Here $A(\boldsymbol{u}, \boldsymbol{\theta})$ and $B(\boldsymbol{u}, \boldsymbol{\theta})$ are $n \times n$ and $p \times p$ unitary matrices respectively. $S(\boldsymbol{u}, \boldsymbol{\theta})$ is a matrix with non-negative singular values on the main diagonal and zeros everywhere else. In fact Assumption A.3 implies that the singular values are all positive.

Due to Assumption A.4 this equality can be extended to a small neighborhood of $(\boldsymbol{u}, \boldsymbol{\theta})$ so that the matrix $A(\boldsymbol{u}, \boldsymbol{\theta})$ is unitary entry-wise continuously differentiable on this neighborhood and $S(\boldsymbol{u}, \boldsymbol{\theta})$ has non-negative entries on diagonal and zeros everywhere else but the entries might no longer be in a decreasing order.

Let $\boldsymbol{\theta}(\boldsymbol{u}) = \arg\min_{\boldsymbol{\theta}} \|\boldsymbol{y} - \boldsymbol{G}(\boldsymbol{u}, \boldsymbol{\theta})\|_2$. Fix $\boldsymbol{y}$ at the observed value and define $\boldsymbol{x}$ through an implicit equation

$$A^\top(\boldsymbol{u}, \boldsymbol{\theta}(\boldsymbol{u}))\{\boldsymbol{y} - \boldsymbol{G}(\boldsymbol{u}, \boldsymbol{\theta}(\boldsymbol{u}))\} - \boldsymbol{x} = 0 \tag{3}$$

Notice that it follows from definition of l^2 projection that $(x_1, \ldots, x_p) = 0$. Furthermore if we set $\boldsymbol{x}_{\complement} = (x_{p+1}, \ldots, x_n)^\top$ we have $\|\boldsymbol{y} - \boldsymbol{G}(\boldsymbol{u}, \boldsymbol{\theta}(\boldsymbol{u}))\|_2 = \|\boldsymbol{x}_{\complement}\|_2$.

We now want to find the density of the random vector $(\boldsymbol{\theta}(\boldsymbol{u}), \boldsymbol{x}_{\complement})$ defined by (3) and evaluated at $\boldsymbol{x}_{\complement} = 0$. By the implicit function theorem there is a neighborhood of $(\boldsymbol{\theta}, 0)$ where $\boldsymbol{u}(\boldsymbol{\theta}, \boldsymbol{x}_{\complement})$ is one to one. The $n \times n$ Jacobian matrix evaluated at $\boldsymbol{x}_{\complement} = 0$ can be directly computed after observing that $\boldsymbol{x}_{\complement} = 0$ implies $\boldsymbol{y} - \boldsymbol{G}(\boldsymbol{u}, \boldsymbol{\theta}(\boldsymbol{u})) = 0$:

$$\left.\frac{\boldsymbol{d}\boldsymbol{u}(\boldsymbol{\theta}, \boldsymbol{x}_{\complement})}{\boldsymbol{d}\boldsymbol{\theta}\boldsymbol{x}_{\complement}}\right|_{\boldsymbol{x}_{\complement}=0} = \left(\frac{\boldsymbol{d}}{\boldsymbol{d}\boldsymbol{u}}\boldsymbol{G}(\boldsymbol{u}, \boldsymbol{\theta})\right)^{-1} A(\boldsymbol{u}, \boldsymbol{\theta}) \left.\left(A^\top(\boldsymbol{u}, \boldsymbol{\theta})\frac{\boldsymbol{d}}{\boldsymbol{d}\boldsymbol{\theta}}\boldsymbol{G}(\boldsymbol{u}, \boldsymbol{\theta}) \ , \ \frac{\boldsymbol{d}\boldsymbol{x}}{\boldsymbol{d}\boldsymbol{x}_{\complement}}\right)\right|_{\boldsymbol{u}=\boldsymbol{G}^{-1}(\boldsymbol{y}, \boldsymbol{\theta})}.$$

Finally denote the first p columns of $A(\boldsymbol{u}, \boldsymbol{\theta})$ by $A_1(\boldsymbol{u}, \boldsymbol{\theta})$. Direct calculation shows that the joint density of $(\boldsymbol{\theta}(\boldsymbol{u}), \boldsymbol{x}_{\complement})$ evaluated at $\boldsymbol{x}_{\complement} = 0$ is

$$h_2(\boldsymbol{\theta}, 0) = f(\boldsymbol{y}|\boldsymbol{\theta}) \left|\det\left(A_1^\top(\boldsymbol{u}, \boldsymbol{\theta})(\frac{\boldsymbol{d}}{\boldsymbol{d}\boldsymbol{\theta}}\boldsymbol{G}(\boldsymbol{u}, \boldsymbol{\theta}))\right)\right|_{\boldsymbol{u}=\boldsymbol{G}^{-1}(\boldsymbol{y}, \boldsymbol{\theta})}.$$

Moreover the properties of singular value decomposition imply that

$$\left|\det\left(A_1^\top(\boldsymbol{u}, \boldsymbol{\theta})(\frac{\boldsymbol{d}}{\boldsymbol{d}\boldsymbol{\theta}}\boldsymbol{G}(\boldsymbol{u}, \boldsymbol{\theta}))\right)\right| = \sqrt{\det\left(\left(\frac{\boldsymbol{d}}{\boldsymbol{d}\boldsymbol{\theta}}\boldsymbol{G}(\boldsymbol{u}, \boldsymbol{\theta})\right)^\top \left(\frac{\boldsymbol{d}}{\boldsymbol{d}\boldsymbol{\theta}}\boldsymbol{G}(\boldsymbol{u}, \boldsymbol{\theta})\right)\right)}.$$

Finally a straightforward calculation using continuity implies that the limiting GFD in (2) is $r(\boldsymbol{\theta}|\boldsymbol{y}) \propto h_2(\boldsymbol{\theta}, 0)$ and the result follows.

Remark A.1. Similar calculation can be done also for the l_1 norm. The minimizer $\boldsymbol{\theta}(\boldsymbol{u})$ of the l_1 norm will have some of its p coordinates of the $\boldsymbol{G}(\boldsymbol{u}, \boldsymbol{\theta}(\boldsymbol{u}))$ exactly equal to some p coordinates of $\boldsymbol{y}$. Therefore we may formulate an equation similar to (3) with A being a row permutation of an identity matrix. The final formula will be similar to the l_∞ norm with an additional term depending on KKT conditions indicating if a particular corner of the l_1 ball associated with $\boldsymbol{y_i}$ and a particular quadrant is feasible as a minimizer of the l_1 norm in (2).

B Assumptions of Theorem 2

Sonderegger and Hannig (2014) prove their version of the Bernstein-von Mises theorem using the l_∞ version of the Jacobian (4). In particular they have $J(\boldsymbol{y}, \boldsymbol{\theta}) = \sum_{\boldsymbol{i}} J_0(\boldsymbol{y_i}, \boldsymbol{\theta})$, where the exact form is given in part (iii) of Theorem 1. We will discuss other Jacobian forms at the end of this section.

We start by reviewing the standard conditions sufficient to prove asymptotic normality of the maximum likelihood estimators (Lehmann and Casella, 1998).

Assumption B.1. There are seven parts:

1. The distributions $P_{\boldsymbol{\theta}}$ are distinct.

2. The set $\{y : f(y|\boldsymbol{\theta}) > 0\}$ is independent of the choice of $\boldsymbol{\theta}$.

3. The data $\boldsymbol{Y} = \{Y_1, \ldots, Y_n\}$ are iid with probability density $f(\cdot|\boldsymbol{\theta})$.

4. There exists an open neighborhood about the true parameter value $\boldsymbol{\theta}_0$ such that all third partial derivatives $\left(\partial^3/\partial\boldsymbol{\theta}_i\partial\boldsymbol{\theta}_j\partial\boldsymbol{\theta}_k\right) f(\boldsymbol{y}|\boldsymbol{\theta})$ exist in the neighborhood, denoted by $B(\boldsymbol{\theta}_0, \delta)$.

5. The first and second derivatives of $L(\boldsymbol{\theta}, y) = \log f(y|\boldsymbol{\theta})$ satisfy

$$E_{\boldsymbol{\theta}}\left[\frac{\partial}{\partial\boldsymbol{\theta}_j} L(\boldsymbol{\theta}, y)\right] = 0$$

   and

$$I_{j,k}(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}}\left[\frac{\partial}{\partial\boldsymbol{\theta}_j} L(\boldsymbol{\theta}, y) \cdot \frac{\partial}{\partial\boldsymbol{\theta}_k} L(\boldsymbol{\theta}, y)\right] = -E_{\boldsymbol{\theta}}\left[\frac{\partial^2}{\partial\boldsymbol{\theta}_j\partial\boldsymbol{\theta}_k} L(\boldsymbol{\theta}, y)\right].$$

6. The information matrix $I(\boldsymbol{\theta})$ is positive definite for all $\boldsymbol{\theta} \in B(\boldsymbol{\theta}_0, \delta)$.

7. There exists functions $M_{jkl}(\boldsymbol{y})$ such that

$$\sup_{\boldsymbol{\theta}\in B(\boldsymbol{\theta}_0,\delta)} \left|\frac{\partial^3}{\partial\boldsymbol{\theta}_j\partial\boldsymbol{\theta}_k\partial\boldsymbol{\theta}_l} L(\boldsymbol{\theta}, y)\right| \leq M_{j,k,l}(y) \quad \text{and} \quad E_{\boldsymbol{\theta}_0} M_{j,k,l}(Y) < \infty.$$

Next we state conditions sufficient for the Bayesian posterior distribution to be close to that of the MLE (van der Vaart, 1998; Ghosh and Ramamoorthi, 2003). The prior $\pi(\boldsymbol{\theta})$ used is the limit of Jacobians from Assumption B.4.

Assumption B.2. Let $L_n(\boldsymbol{\theta}) = \sum L(\boldsymbol{\theta}, Y_i)$.

1. For any $\delta > 0$ there exists $\epsilon > 0$ such that

$$P_{\boldsymbol{\theta}_0} \left\{ \sup_{\boldsymbol{\theta} \notin B(\boldsymbol{\theta}_0, \delta)} \frac{1}{n} \left(L_n(\boldsymbol{\theta}) - L_n(\boldsymbol{\theta}_0) \right) \leq -\epsilon \right\} \to 1.$$

2. $\pi\left(\boldsymbol{\theta}\right)$ is positive at $\boldsymbol{\theta}_0$.

Finally we state assumptions on the Jacobian function.

Assumption B.3. For any $\delta > 0$

$$\inf_{\boldsymbol{\theta} \notin B(\boldsymbol{\theta}_0, \delta)} \frac{\min_{\boldsymbol{i}=(i_1,\dots,i_p)} L(\boldsymbol{\theta}, \boldsymbol{Y}_{\boldsymbol{i}})}{|L_n(\boldsymbol{\theta}) - L_n(\boldsymbol{\theta}_0)|} \xrightarrow{P_{\boldsymbol{\theta}_0}} 0,$$

where $L_n\left(\boldsymbol{\theta}\right) = \sum_{i=1}^{n} \log f\left(x_i | \boldsymbol{\theta}\right)$ and $B\left(\boldsymbol{\theta}_0, \delta\right)$ is a neighborhood of diameter δ centered at $\boldsymbol{\theta}_0$.

Assumption B.4. There is a normalization c_n so that the Jacobian function $c_n^{-1} J\left(\boldsymbol{Y}, \boldsymbol{\theta}\right) \overset{a.s.}{\to} \pi\left(\boldsymbol{\theta}\right)$ as $n \to \infty$ uniformly on compacts in $\boldsymbol{\theta}$.

Assumption B.4 can be verified in the one-parameter case using the classical Uniform Strong Law of Large Numbers (van der Vaart, 1998; Ghosh and Ramamoorthi, 2003).

In the multi-parameter case Jacobian function $J(\boldsymbol{Y}, \boldsymbol{\theta}) = \sum_{\boldsymbol{i}} J_0(\boldsymbol{Y}_{\boldsymbol{i}}, \boldsymbol{\theta})$ is a U-statistic and the uniform convergence follows from Yeo and Johnson (2001) with $c_n = \binom{n}{p}$. In particular, Assumption B.4 is implied by Assumption B.5 below.

Assumption B.5. Let $\boldsymbol{j}$ be a multi-index with values in $\{1, 2, \dots p\}$ and denote a vector $\boldsymbol{y}_{\boldsymbol{j}} = (y_{i_1}, \dots, y_{i_k})$. Next define the complement multi-index $\boldsymbol{j}^{\complement} = \{i, i \notin \boldsymbol{j}\}$ and its corresponding vector $\boldsymbol{y}_{\boldsymbol{j}^{\complement}}$.

1. There exists a symmetric function $g(\cdot)$ integrable with respect to $P_{\boldsymbol{\theta}_0}$, and compact space $\bar{B}(\boldsymbol{\theta}_0,\delta)$ such that for $\boldsymbol{\theta} \in \bar{B}(\boldsymbol{\theta}_0,\delta)$ and $\boldsymbol{y} \in \mathbb{R}^p$ then $|J_0(\boldsymbol{y};\boldsymbol{\theta})| \le g(\boldsymbol{y})$.

2. There exists a sequence of measurable sets S_M^p such that

$$P\left(\mathbb{R}^p - \cup_{M=1}^{\infty} S_M^p\right) = 0,$$

3. For each M and for all $\boldsymbol{j}$,

$$J_{\boldsymbol{j}}\left(\boldsymbol{y}_{\boldsymbol{j}};\boldsymbol{\theta}\right) = E_{\boldsymbol{\theta}_0}\left[J_0\left(\boldsymbol{y}_{\boldsymbol{j}},\boldsymbol{Y}_{\boldsymbol{j}^{\complement}};\boldsymbol{\theta}\right)\right].$$

   is equicontinuous in $\boldsymbol{\theta} \in \bar{B}(\boldsymbol{\theta}_0,\delta)$ for $\{\boldsymbol{y}_{\boldsymbol{j}}\} \in S_M^{\boldsymbol{j}}$ where $S_M^p = S_M^{\boldsymbol{j}} \times S_M^{\boldsymbol{j}^{\complement}}$.

B.1 Extension to More General Jacobians

Notice that when $\boldsymbol{Y} = (Y_1,\ldots,Y_n)^\top$ are i.i.d., the l_∞ Jacobian satisfies $\int J_0(\boldsymbol{y}_{\boldsymbol{i}},\boldsymbol{\theta}) f(\boldsymbol{y}_{\boldsymbol{i}}|\boldsymbol{\theta}) d\boldsymbol{\theta} = 1$ for all $\boldsymbol{i} = (i_1 < \cdots < i_p)$.

For other version of the Jacobian we need to additionally assume that there are $J(\boldsymbol{y},\boldsymbol{\theta}) = \sum_{\boldsymbol{i}} \tilde{J}_{\boldsymbol{i}}(\boldsymbol{y},\boldsymbol{\theta})$, and a constant C so that $\int \tilde{J}_{\boldsymbol{i}}(\boldsymbol{y},\boldsymbol{\theta}) f(\boldsymbol{y}_{\boldsymbol{i}}|\boldsymbol{\theta}) d\boldsymbol{\theta} \le C$ for all $\boldsymbol{i} = (i_1 < \cdots < i_{p'})$ and all n big enough.

Finally, we remark that Assumption B.4 becomes relatively easier to verify when considering the l_2 Jacobian from part (ii) of Theorem 1, as one can use the Uniform Law of Large Numbers instead of uniform convergence of U-statistics.

C Proof of Theorem 3

Proof. Using Skorokhod's representation theorem we can assume that there is a version of data so that $t_n(\boldsymbol{Y}_n) \to \boldsymbol{T}$ almost surely. The theorem is then proved in three steps.

We need to compute

$$P(\boldsymbol{\theta}_{n,0} \in C_n(\boldsymbol{Y}_n)) = P(\boldsymbol{\xi}_0 \in \Xi_n(\boldsymbol{Y}_n)) \ge P(\boldsymbol{\xi}_0 \in C(\boldsymbol{T})) - P\left(\boldsymbol{\xi}_0 \in C(\boldsymbol{T}) \setminus \bigcap_{k=n}^{\infty} \boldsymbol{\Xi}_k(C_k(\boldsymbol{Y}_k))\right).$$

Since the set $C(\boldsymbol{t}) \setminus \bigcap_{k=m}^{\infty} \Xi_n(C_n(\boldsymbol{Y}_n))$ shrinks monotonically to $\emptyset$, $\boldsymbol{\xi}_0$ will be excluded eventually almost surely and the last probability in the equation above goes to zero. Analogously

$$P(\boldsymbol{\theta}_{n,0} \in C_n(\boldsymbol{Y}_n)) \leq P(\boldsymbol{\xi}_0 \in C(\boldsymbol{T})) + P\left(\boldsymbol{\xi}_0 \in \bigcup_{k=n}^{\infty} \boldsymbol{\Xi}_k(C_k(\boldsymbol{Y}_k)) \setminus C(\boldsymbol{T})\right).$$

By combining these two inequalities we get $P(\boldsymbol{\theta}_{n,0} \in C_n(\boldsymbol{Y}_n)) \to P(\boldsymbol{\xi}_0 \in C(\boldsymbol{T}))$.

Next, define $V_{\boldsymbol{t}}(\boldsymbol{\xi}) = \{\boldsymbol{v} : \boldsymbol{t} = \boldsymbol{H}(\boldsymbol{v}, \boldsymbol{\xi})\}$. Notice that $V_{\boldsymbol{t}}(C(\boldsymbol{t})) = \mathcal{V}_{\boldsymbol{t}_2}$ defined in the statement of the Theorem 3. Also by invertibility $V_{\boldsymbol{T}}(\boldsymbol{\xi}_0)$ has the same distribution as $\boldsymbol{V}$ in the limiting data generating equation. Recall that the limiting GFD $R_{\boldsymbol{t}}$ is the conditional distribution $Q_{\boldsymbol{t}_1}(\boldsymbol{V}_1^\star) \mid \boldsymbol{H}_2(\boldsymbol{V}_2^\star) = \boldsymbol{t}_2$, where $Q_{\boldsymbol{t}_1}(\boldsymbol{v}_1) = \boldsymbol{\xi}$ is the solution of $\boldsymbol{t}_1 = \boldsymbol{H}_1(\boldsymbol{v}_1, \boldsymbol{\xi})$. Consequently,

$$P(\boldsymbol{\xi}_0 \in C(\boldsymbol{T})) = P(V_{\boldsymbol{T}}(\boldsymbol{\xi}_0) \in \mathcal{V}_{\boldsymbol{T}_2}) = EP((\boldsymbol{V}_1, \boldsymbol{V}_2) \in \mathcal{V}_{\boldsymbol{T}_2} | H_2(\boldsymbol{V}_2) = \boldsymbol{T}_2) = E[R_{\boldsymbol{T}}(C(\boldsymbol{T}))],$$

where the second equality follows from the fact that conditionally on $\boldsymbol{T}_2 = \boldsymbol{t}_2$ the set $\mathcal{V}_{\boldsymbol{t}_2}$ is a fixed (non-random) set.

We conclude by showing that $R_{\boldsymbol{t}}(C(\boldsymbol{t})) = \alpha$ for almost all $\boldsymbol{T} = \boldsymbol{t}$. Denote $A_m = \bigcap_m^{\infty} \Xi_n(C_n(\boldsymbol{y}_n))$ and $B_m = \bigcup_m^{\infty} \Xi_n(C_n(\boldsymbol{y}_n))$. By our assumptions $R_{\boldsymbol{t}}(\partial A_m) = R_{\boldsymbol{t}}(\partial B_m) = 0$ and consequently we have $\alpha \geq \lim_{n\to\infty} R_{n,\boldsymbol{y}_n} \Xi_n^{-1}(A_m) = R_{\boldsymbol{t}}(A_m) \to R_{\boldsymbol{t}}(C(\boldsymbol{t}))$ and $\alpha \leq \lim_{n\to\infty} R_{n,\boldsymbol{y}_n} \Xi_n^{-1}(B_m) = R_{\boldsymbol{t}}(B_m) \to R_{\boldsymbol{t}}(C(\boldsymbol{t}))$ by continuity of measure. The statement of the theorem follows. $\square$

D Proof of Theorem 4

We will study GFD defined for a finite collection of models $\mathcal{M}$. Recall that the data generating equation is

$$\boldsymbol{Y} = \boldsymbol{G}(M, \boldsymbol{\theta}_M, \boldsymbol{U}), \qquad M \in \mathcal{M},\ \boldsymbol{\theta}_M \in \boldsymbol{\Theta}_M,$$

where $\boldsymbol{y}$ is the observations, M is the model considered, $\boldsymbol{\theta}_M$ are the parameters associated with model M, and $\boldsymbol{U}$ is a random vector of with fully known distribution independent of any parameters. To derive GFD in this context we will apply definition in (2).

We start by stating an assumption closely related to identifiability.

Assumption D.1. For any two models $M_1 \neq M_2 \in \mathcal{M}$,

$$P\left(\bigcap_{i=1,2}\{\min_{\boldsymbol{\theta}_{M_i},\sigma^2}\|\boldsymbol{y}-\boldsymbol{G}(M_i,\boldsymbol{\theta}_{M_i},\boldsymbol{U})\|\le\epsilon\}\right)$$

$$= o\left(\max_{i=1,2} P(\min_{\boldsymbol{\theta}_{M_i},\sigma^2}\|\boldsymbol{y}-\boldsymbol{G}(M_i,\boldsymbol{\theta}_{M_i},\boldsymbol{U})\|\le\epsilon)\right), \quad \epsilon\to 0. \quad (4)$$

A simple calculation applying the inclusion and exclusion formula to (2) gives the following result.

Lemma D.1. *Under Assumption D.1 the marginal fiducial distribution for each $M \in \mathcal{M}$ is the limit, as $\epsilon \to 0$, of the conditional probabilities*

$$r(M|\boldsymbol{y}) = \lim_{\epsilon\to 0}\frac{P(\min_{\boldsymbol{\theta}_M,\sigma^2}\|\boldsymbol{y}-\boldsymbol{G}(M,\boldsymbol{\theta}_M,\boldsymbol{U})\|\le\epsilon)}{x}\sum_{M'\in\mathcal{M}}P(\min_{\boldsymbol{\theta}_{M'},\sigma^2}\|\boldsymbol{y}-\boldsymbol{G}(M',\boldsymbol{\theta}_{M'},\boldsymbol{U})\|\le\epsilon). \quad (5)$$

We are now ready to prove Theorem 4. In the rest of this section we also suppose the assumptions of Theorem 1 hold. First notice that the invertibility implies $|M| \le n$, where $|M|$ is the number of parameters in M. More importantly, as $\epsilon \to 0$

$$P\left(\min_{\boldsymbol{\theta}_M,\sigma^2}\|\boldsymbol{y}-\boldsymbol{G}(M,\boldsymbol{\theta}_M,\boldsymbol{U})\|_\infty\le\epsilon\right)\sim C_M(\boldsymbol{y})\epsilon^{\min(0,n-|M|)},$$

where $C_M(\boldsymbol{y}) = \int_{\boldsymbol{\Theta}_M} f_M(\boldsymbol{y},\boldsymbol{\theta}_M)J_M(\boldsymbol{y},\boldsymbol{\theta}_M)\,d\boldsymbol{\theta}_M$. Consequently the GFD in (5) assigns positive probability only to the largest model. To solve this issue we augment for each model the data generating equation $\boldsymbol{Y} = \boldsymbol{G}(M,\boldsymbol{\theta}_M,\boldsymbol{U})$ by

$$p_k = P_k, \quad k = 1,\ldots,|M|,$$

where P_i are i.i.d. continuous random variables with $f_P(0) = q$ independent of $\boldsymbol{U}$, and q is a constant determined by the penalty. Since these extra generating equations are fully synthetic we can set the observed value to $p_i = 0$. For the augmented data generating equation we get

$$P\left(\min_{\boldsymbol{\theta}_M,\sigma^2}\|\boldsymbol{y}-\boldsymbol{G}(M,\boldsymbol{\theta}_M,\boldsymbol{U})\|_\infty\le\epsilon,\max_{i=1,\ldots,|M|}|P_i|\le\epsilon\right)\sim C_M(\boldsymbol{y})q^{|M|}\epsilon^n.$$

The statement of Theorem 4 follows.

To conclude we remark that if the original data generating equation satisfied the identifiability assumption, so does the augmented data generating equation. To see this, notice that if $|M_1| \leq |M_2|$ then the left-hand-side of (4) is multiplied by a factor of order $\epsilon^{|M_2|}$ while the terms on the right-hand-side of (4) are multiplied only by a factor of $O(\epsilon^{|M_2|})$.

E Assumptions for Theorem 5

Assumption E.1. This assumption has four parts:

1. Assume that $\mathbb{R}$ is partitioned into fixed intervals

$$(-\infty, a_1], (a_1, a_2], \ldots, (a_k, \infty)$$

denoting $a_0 = -\infty,\ a_{k+1} = \infty$.

The values of Y_i are observed only up to the resolution of the grid. In other words, we do not observe the realized value y_i itself, only which of the intervals it falls into; i.e., we observe $\boldsymbol{k} = (k_1, \ldots, k_n)$ so that $y_i \in (a_{k_i}, a_{k_i+1}]$ or equivalently $\boldsymbol{y} \in (\boldsymbol{a_k}, \boldsymbol{a_{k+1}}]$ with $\boldsymbol{a_k} = (a_{k_1}, \ldots, a_{k_n})$.

2. Assume that $k \geq p$. For all $j = 0, \ldots, k$ and all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ we have $P_{\boldsymbol{\theta}}\left(Y \in (a_j, a_{j+1})\right) > 0$ and $P_{\boldsymbol{\theta}}\left(Y = a_j\right) = 0$.

3. Assume $F(y|\boldsymbol{\theta})$ is continuously differentiable in $\boldsymbol{\theta}$ for all $y \in \{a_1, \ldots, a_k\}$; i.e., all p first order partial derivatives are continuous.

4. For all $\boldsymbol{j} = (j_1 < \cdots < j_p) \subset \{1, \ldots, k\}$ and $u_1 < \cdots < u_p$ there is a unique solution $\boldsymbol{\theta}$ of $(F(a_{j_1}|\boldsymbol{\theta}), \ldots, F(a_{j_p}|\boldsymbol{\theta})) = (u_1, \ldots, u_p)$, and the Jacobian

$$\det\left(\frac{\boldsymbol{d}(F(a_{j_1}|\boldsymbol{\theta}), \ldots, F(a_{j_p}|\boldsymbol{\theta}))}{\boldsymbol{d\theta}}\right) \neq 0.$$

References

Ghosh, J. K. and Ramamoorthi, R. V. (2003) *Bayesian Nonparametrics*. Spring-Verlag.

Hannig, J. (2013) Generalized Fiducial Inference via Discretization. *Statistica Sinica*, **23**, 489–514.

Lehmann, E. L. and Casella, G. (1998) *Theory of point estimation*. New York: Springer.

Sonderegger, D. and Hannig, J. (2014) Fiducial theory for free-knot splines. In *Contemporary Developments in Statistical Theory, a Festschrift in honor of Professor Hira L. Koul* (eds. S. Lahiri, A. Schick, A. SenGupta and T. Sriram), 155 – 189. Springer.

van der Vaart, A. W. (1998) *Asymptotic statistics*, vol. 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge: Cambridge University Press.

Yeo, I. K. and Johnson, R. A. (2001) A uniform strong law of large numbers for U-statistics with application to transforming to near symmetry. *Statistics and Probability Letters*, **51**, 63–69.